DOCUMENT RESUME

ED 147 275 95 SP 011 853

AUTHOR	Crawford, W. John; And Others
TITLE	Error Rates and Question Difficulty Related to Elementary Children's Learning. Report No. 75-8.
INSTITUTION	Texas Univ., Austin. Research and Development Center for Teacher Education.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	75
CONTRACT	OEC-6-10-108
NOTE	19p.; Paper presented at the Annual Meeting of the American Educational Research Asscciation (Washington, D.C., April 1975); For related documents, see SP011 799, 852, and 855
EDRS PRICE	MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS	Academic Achievement; *Complexity Level; *Effective Teaching; *Elementary Education; *Learning Processes; Performance Factors; *Questicning Techniques; *Socioeconomic Influences

ABSTRACT

Elementary school teachers chosen for their effectiveness in consistently producing achievement gains in their pupils were observed to determine how difficult a teacher's questions should be for students to learn the maximum amount of material. The collected data, correlated with residualized student achievement gain scores (averaged over four years), produced a pattern of different optimal error rates and different optimal question difficulty levels for high socioeconomic status as opposed to low socioeconomic status schools. These data are presented and their implications for teaching strategies discussed. (Author/MJB)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED147275

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: SP TM
In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

"Error Rates and Question Difficulty Related to Elementary Children's Learning"[1]

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

W. John Crawford,
Carol E. King,
Jere E. Brophy,
and
Carolyn M. Evertson
The University of Texas at Austin

Report No. 75-8

[1]Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C., April, 1975.

SP011 853

This project was supported by the National Institute of Education Contract OEC 6-10-108, Research and Development Center for Teacher Education. The opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement by that office should be inferred.

The present paper was one of four presented at a "related papers" session at the 1975 annual meeting of the American Educational Research Association, entitled "Teacher Behavior and Student Outcomes: Texas Teacher Effectiveness Study."

Papers in the session included:

1. "Teacher Behaviors Related to Learning by Low vs. High Socio-economic Status Early Elementary Students," by Jere E. Brophy.
2. "Relationship of Teacher Praise and Criticism to Student Outcomes," by Carolyn M. Evertson.
3. "Teacher Feedback to Children's Answers: Process-Product Relationships," by Linda L. Mahaffey, Jere E. Brophy, and Carolyn M. Evertson.
4. "Error Rates and Question Difficulty Related to Elementary Children's Learning," by W. John Crawford, Carol E. King, Jere E. Brophy, and Carolyn M. Evertson.

This paper and the others presented in this session will all present data from the Texas Teacher Effectiveness Project, a two-year observational study of second and third grade teachers designed to identify relationships between teacher characteristics and student learning. This paper will deal with some of the relationships which were found between several questioning difficulty variables and student outcome data. Before getting to these data, however, we will first briefly describe the study as a whole for those of you who are unfamiliar with it.

The study was a two-year, replicated, naturalistic observational study involving collection of classroom process data through both high and low inference instruments and presage data through questionnaires and interviews. The teacher effectiveness criteria were residual student gain scores on three language arts and two mathematics subtests of the Metropolitan Achievement Test battery. The teachers included in this sample had been teaching at either second or third grade for a minimum of four years when the study began, and they were among the most consistent teachers in producing student learning gains across time (school years), sex of student, and subtest. Residual gains for each student were computed within each year, within each subtest, within sex, and within Title I versus Non-Title I schools (because different tests were used in these two different kinds of schools). These residual gain scores for individual students were then summed and averaged to produce a mean residual gain for each teacher's class. This was done for three consecutive years for a total of 165 teachers. From these, those who were

most consistent across the five subtests, the two sexes, and the three years of data were selected for study. This was done because it was felt that the teachers most consistent in producing student learning gains would be most likely to behave consistently in the classroom and thus yield clear and interpretable process-product relationships. For detailed reports of the rationale and statistics involved in sample selection, as well as for details of the data discussed in the present report, see the bibliography at the end of the paper.

Thirty-one teachers were observed in the first year for about 10 hours each, and 28 teachers were observed in the second year for about 30 hours each. Nineteen teachers were observed in both years, since we lost 12 of the first year teachers and added nine more to the sample for the second year. The low-inference system used was based on the Brophy-Good Dyadic Interaction Observation System (Brophy and Good, 1970), although it was considerably expanded to take into account numerous contextual variables and other interaction variables which we thought were important, and also to accommodate some of the variables that Kounin (1970) had stressed in his research on classroom management. Data also were collected through high-inference ratings of the teachers made by the observers at the end of the year, and through interviews and questionnaires designed to get at information concerning teacher beliefs, attitudes, and practices which could not be observed in the classroom, such as preparation and test construction.

An attempt was made to collect affective data concerning student perceptions of the teacher and the school, but this proved to be unproductive; the students did not give valid responses, apparently because of developmental

immaturity. Thus, it should be kept in mind that the only criteria of teacher effectiveness in the study, except when process variables in the affective sphere are mentioned, was teacher ability to produce student learning gains on standardized achievement tests. This is important, because many investigators have found that teacher behavior that fosters learning gains is unrelated or even negatively related to teacher behavior that fosters popularity with students, and vice versa.

Each presage or process variable was analyzed for linear correlations and for non-linear relationships with student outcome criteria. These analyses produced a large volume of data, some of which supports the existing literature and some of which seems to contradict it. Discussion of some of these contradictions and corroborations will be given in this paper and the others presented in this session. The present report concentrates on relationships between the difficulty level (error rate) of the teachers' questioning and student learning gains. We will also discuss how the optimal error rate seems to vary with certain affective variables, socio-economic status, and the ability of the student.

Questioning: How Difficult or How Easy Should It Be?

The general issue addressed here is how difficult a teacher's questions should be in order for his or her students to learn the maximum amount of material. That is, should teachers ask hard-to-answer questions, working on the assumption that children learn primarily from their mistakes, or should they teach with non-challenging, easy-to-answer questions, thus avoiding most frustration and/or negative reinforcement which might come from not experiencing enough success?

The very idea of individualizing instruction implies that there is a certain level of difficulty or content specificity that is optimal for each student. This idea has been propounded by writers with views as different as those of Hunt (1961), who describes "the problem of the match" (the need for teachers to match materials and assignments to the present interests and abilities of the student), and Ausubel (1963), who has pointed out that the most important determinants of student learning are the student's present level of ability and his or her knowledge in the subject matter area (implying that learning will proceed optimally if new input is matched to present structures). The statements of these writers, the concept of individualization, and numerous related concepts all imply a curvilinear relationship between the difficulty level of instruction and the value of the learning experience. Learning would seem to be most optimal when presented at the optimal difficulty level. If the material is either too easy or too difficult, learning will not proceed as well.

This fundamental principle showed itself many different ways in our data. Perhaps the most obvious area was in the percentages of teacher questions which were answered correctly by the students (which in turn were related to the difficulty levels of the questions that the teachers asked). Linear correlations for both years of the study showed that the percentage of correct answers in low SES schools (about 70%) was generally positively related to student learning gains, while the percentage of correct answers in the high SES schools (about 80%) was found to be negatively related to gains. Upon examining non-linear analyses, we found that when a relationship between a "difficulty level" variable and a student learning gain measure was

significantly non-linear, the shape of the curve was that of an inverted "U". This suggested that there indeed is an optimal level of question difficulty, and that questioning rates significantly below or above this optimal rate will result in less than maximal learning.

Several things are noteworthy about these findings. First, the curvilinear relationships confirm the general ideas underlying individualization, particularly the idea that there is an optimal level of difficulty which will produce more learning than a level which is either easier or more difficult. Second, the SES difference in correlational patterns provides an indication of the nature of the kind of individualization that is required. Since SES is a "proxy" variable which stands for a complex of ability, achievement, and motivation, extrapolation from this finding suggests that low SES children (and, by implication, less motivated and less able children of any SES level) learn best by having smaller "chunks" taught to them per unit of time; that is, by having material taught to them redundantly to the point of overlearning, proceeding in small steps so that they can master the material without undue cognitive strain. In contrast, somewhat more confident and able children can cover the same material more quickly, and furthermore will learn more if challenged with slightly more difficult questions and assignments. Questions which challenge high achieving students generate more wrong answers, but they also prompt more thinking and reasoning and therefore more learning. For low achieving students, the greater danger may lie in asking questions that are too difficult and frustrating to allow maximimal learning to occur.

Another item worth noting concerning these findings is their bearing on the "errorless learning" hypothesis. Errorless learning advocates, who include most writers of programmed texts, assume that learning is optimal when it proceeds errorlessly. Taken literally, this would mean that learning would be optimal when the student answered 100% of questions or assignment items correctly. However, our non-linear regression data dispute this. The data do indicate that learning proceeds most smoothly when the material is relatively easy to assimilate (optimal levels were up around 75% of questions answered correctly, rather than down at lower levels which would indicate more difficult material). However, the percentages are still considerably different from 100%. What this may mean is that, although learning does proceed most smoothly when there is no undue cognitive strain, learning must also involve some new challenge and a certain amount of failure or confusion if there is going to be any significant forward progress.

While errorless learning advocates would posit a 100% success rate as optimal, social scientists concerned with achievement motivation might posit a 50% success rate as optimal. This idea is based on several laboratory experiments which have indicated that achievement motivation is strongest when the probability of succeeding is 50%. However, taking these data and attempting to apply them directly to the problem of the difficulty level of classroom material is a classic example of attempting to generalize laboratory data to naturalistic situations without taking other relevant factors into account.

First of all, these data have come almost exclusively from game-like experimental situations, such as horseshoes, ring toss, and building block towers. Most of the activities involved were physical skill activities involving little or no cognitive work. Thus, while it may be true that physical skill activities are most enjoyable and in fact engender the most achievement motivation when they are at about a 50% difficulty level, it appears that cognitive activities which are at a 50% difficulty level are too difficult and produce too much cognitive strain for most people.

Also, the data from achievement motivation experiments involve relationships between difficulty level and achievement motivation (the desire to achieve), as contrasted with the relationships between difficulty level and measures of achievement per se. That is, they concern the relationship between difficulty level and the person's desire to achieve on the task, not his actual achievement on the task. Other studies of motivation have suggested that motivation itself is curvilinearly related to achievement, so that people who have optimally high levels of motivation do better than people who are either under- or over-motivated.

Undoubtedly, there is no one exact "magic percentage" or optimal success rate, since we would expect changes depending upon individual students' interest in and aptitude for the subject matter, as well as on general differences in ability and achievement. A summary statement on this issue would be that new input is at the optimal difficulty level when the student makes as much progress as he or she can without becoming frustrated due to cognitive strain.

Another cluster of variables dealing with the topic of the teacher's questioning strategies includes frequencies of process, product, and choice questions which the teacher asked in public response opportunities. Process questions were the most difficult type of questions; product, the next; and choice questions, the simplest type. The findings concerning level of difficulty of teachers' questions as measured by these variables were generally mixed and failed to fit any discernible pattern. These mixed findings are partially accounted for by the fact that the frequency of occurrence of process questions, as opposed to "fact related" questions (product and choice), was very low; i.e., the frequency distributions showed mostly product and choice questions, with very few process questions being asked in the second and third grades. At this early elementary level, the emphasis is on the acquisition of "tool skills," and this is most easily facilitated by factual questions which require only short responses, not by long, involved process questions. The authors currently are conducting a study in the seventh and eighth grades which should yield data shedding more light on the role of process questions and the relationships between process, product, and choice questions and student learning gains.

Assignment Difficulty

A high inference variable, observer ratings of assignment difficulty, showed an interesting contrast by SES in patterns of correlations with student learning gains. In low SES classrooms, negative correlations were obtained between ratings of assignments as being too hard and measures of student learning gains. In contrast, ratings of assignments as being too easy correlated negatively with learning gains in high SES schools. Here again is evidence that it was important for the low SES teachers to concentrate

on teaching to the point of overlearning, but important for high SES teachers to move at a faster pace and to provide more variety and challenge. It was especially important for teachers in low SES schools to err (if they erred at all) by overdwelling on work too long rather than by moving on too quickly. It seemed that students in low SES schools tended to give up when the work was too hard, withdrawing into daydreams or copying from their neighbors. This may be another aspect of locus of control; these students apparently did not think that they could master the material (students who did think so would go to the teacher for help and persist in trying to solve their problem rather than withdrawing from the task or simply copying).

Some Analogous Experimental Data

Crawford (Note 1) conducted two experimental studies designed to follow up these correlational results through systematic variation of the difficulty level of programmed learning materials, using college students as subjects. He did find a direct linear relationship between easiness of materials and degree of mastery, just as errorless learning advocates would predict. However, he also found a trend of increasing difficulty level being associated with higher gains for high GPA subjects (as opposed to the ineffectiveness of these difficult materials when used by low GPA subjects). That is, when he made within treatment-group comparisons of low GPA vs. high GPA subjects, he discovered a definite trend of increasing differences between the two ability groups (with high GPA subjects doing better on the dependent measure of amount learned) with an increase in the error rates of the teaching materials. With the more difficult sets of materials, the high GPA subjects did quite well, while the performance (assimilation of information covered in the

materials) of the low GPA subjects markedly fell off as difficulty level increased. The negative relationship found between error rate and amount of learning, as opposed to the "inverted-U" curvilinear relationships found in some of the second and third grade data, could have resulted from differing modes of teaching. The elementary school data were taken from teachers asking questions and providing feedback during public response opportunities in the classroom, while Crawford's data came from college students working independently with programmed materials. The different results could be connected with a cognitive development trend--the elementary school children were mostly concrete operational while the college students were in formal operations. Or, the reason could be the one alluded to previously in the Ausubel (1963) citation: that the student's present level of ability and knowledge in a subject matter area are important determinants of student learning. (Crawford's subjects initially were completely unfamiliar with the topic of the materials; thus, in order to match the new input to present structures and maximize learning gains, the materials would have to proceed in small steps, with low error rates. This probably would <u>not</u> be the case in the second and third grade situation.)

It may be somewhat difficult to draw clear conclusions and implications here; however, the two data sets (the second and third grade correlational data and the Crawford experimental data), in combination with other considerations, seem to mutually support the idea of a complex aptitude-treatment interaction involving various internal predispositions of subjects, difficulty level, and mode of presentation of questions intended to teach the material. Crawford's "GPA" variable seemed to operate similarly to the "SES" variable in the second and third grade study.

Conclusions

Regardless of age or grade levels, it seems intuitively obvious that learning assignments must be easy (i.e., the student probably can complete them without help) in situations where the student will be expected to work alone with little or no opportunity to get feedback and/or help. In contrast, when the teacher or other resources are available to present material and interact with students (either in group lessons or in individual interactions concerning work assignments), it is possible and apparently optimal for the teacher to present new, more challenging, and more difficult material. The same students might not learn it clearly if working completely on their own, but the teacher's presence makes it possible for them to ask questions, to get feedback, and generally to have the material presented to them in ways that they can understand.

A related principle to the one concerning availability of the teacher to provide feedback concerns the ability and knowledge levels of the learners. The higher these are, the more difficult and challenging the task can become without losing effectiveness. Conversely, less able or knowledgeable students will need the material presented in smaller chunks and with greater redundance. These generalizations seem consistent with both our own correlational data from elementary classrooms and with the Crawford (Note 1) experimental data from college students working independently.

Many of our findings concerning difficulty level of questions and assignments fit in with this general interpretation. Thus, for example, the more successful teachers in the low SES schools moved at a relatively slow pace and spent much time teaching and reteaching the basic fundamentals

of reading, writing, and arithmetic. Most assignments were at the level of skill practice and factual memory. In contrast, the more successful high SES teachers blended in problem solving and other high level activities which involved application of the skills as well as mastery of the skills themselves. They also worked at a faster pace and introduced more variety in their teaching methods and materials. For them, providing the students with interesting variety and stimulating challenges in their assignments was more important than seeing that they got continual and cumulative practice in mastering the fundamentals (since many already had mastered the fundamentals, at least by the third grade).

When a student is having difficulty with even simple assignments involving basic tool skills, it is important to teach him in very small steps to the point of overlearning and to drill him on fundamentals. However, as he masters these fundamentals, it is more important to gradually phase in problem solving, faster pacing, more variety, and other aspects of teaching that are more optimal for the kind of student that he has now become.

Thus, we see the task of matching instruction to student needs as a complex, multivariate task involving such variables as: mode of presentation of material, the developmental level of the learner, the general ability and achievement motivation level of the learner, the familiarity of the materials, and the difficulty of the questions designed to teach the material. It is hoped that, by application of data discussed here, some insight may be lent to the formulation of a model of instruction which would maximize learning for all students.

Reference Note

Crawford, W. J. An examination of the effect of error rates and grade point averages on learning gains. (Res. Rep. 75-10). Austin, Texas: Research and Development Center for Teacher Education, 1975. Paper presented at the annual convention of the Southwest Psychological Association held in Houston, Texas, 1975.

References

Ausubel, D. The psychology of meaningful verbal learning. New York: Grune and Stratton, 1963.

Brophy, J. and Good, T. The Brophy-Good dyadic interaction system. In A. Simon and E. Boyer (Eds.). Mirrors for behavior: An anthology of observation instruments continued, 1970 supplement, Volume A. Philadelphia: Research for Better Schools, Inc. 1970.

Hunt, J. Intelligence and experience. New York: Ronald Press, 1961.

Kounin, J. Discipline and group management in classrooms. New York: Holt, Rinehart and Winston, 1970.

Bibliography

Complete presentations of the research design, data collection procedures, and data analyses are available in the sources listed below.

Brophy, J. Stability in teacher effectiveness. (Res. Rep. 77). Austin, Texas: Research and Development Center for Teacher Education, 1972 (ERIC No. ED 066 438).

Brophy, J. Stability of teacher effectiveness. American Educational Research Journal, 1973, 10, 245-252.

Brophy, J. and Evertson, C. Low-inference observational coding measures and teacher effectiveness. Catalog of Selected Documents in Psychology, 1973 (a), 3, 97.

Brophy, J. and Evertson, C. Appendix to first-year data of Texas teacher effectiveness project: Complex relationships between teacher process variables and student outcome measures. Catalog of Selected Documents in Psychology, 1973 (b), 3, 137.

Brophy, J. and Evertson, C. The Texas teacher effectiveness project: Summary report. (Res. Rep. 74-4). Austin, Texas: Research and Development Center for Teacher Educationa, 1974 (b).

Brophy, J. and Evertson, C. The Texas teacher effectiveness project: Presentation of non-linear relationships and summary discussion. (Res. Rep. 74-6). Austin, Texas: Research and Development Center for Teacher Education, 1974 (b).

Evertson, C. and Brophy, J. High-inference behavioral ratings as correlates of teaching effectiveness. Catalog of Selected Documents in Psychology, 1973, 3, 97.

Evertson, C. and Brophy, J. Texas teacher effectiveness project: Questionnaire and interview data. (Res. Rep. 74-5). Austin, Texas: Research and Development Center for Teacher Education, 1974.

Veldman, D. and Brophy, J. Measuring teacher effects on pupil achievement. Journal of Educational Psychology, 1974, 66, 319-324.